Contextual Dynamics of Group-Based Sharing Decisions

Simon Jones and Eamonn O'Neill
Department of Computer Science,
University of Bath,
Bath, BA2 7AY, UK
s.jones2@bath.ac.uk, eamonn@cs.bath.ac.uk

ABSTRACT
In this paper we investigate how decisions made while using a granular access control mechanism for sharing photographs are influenced by contextual factors and properties relating to the identities of contacts. We develop analytical models using logistic regression to understand relationships between variables that affect sharing decisions. We also investigate how predefined, static groups for privacy control cope with the challenge of sharing large amounts of content associated with numerous different contexts, and test whether they need to be adjusted to suit particular contexts.

AUTHOR KEYWORDS
Privacy, Social Media, Photo Sharing, Context-Awareness.

ACM CLASSIFICATION KEYWORDS
H5.3. Group and Organization Interfaces.

GENERAL TERMS
Human Factors, Measurement, Theory

INTRODUCTION

Disclosing personal content on social networking services can expose sensitive information about users. These services typically allow users to create connections to 'friends' such that this content can be shared amongst them and restricted from the wider public. However, these connections rarely distinguish between different types of relationship. Even within a network of 'friends', users may wish to manage the sharing of information and content with different people based on their differing relationships.

Previous work, e.g. [8, 13], has investigated the potential of automating the creation of privacy-based groups for controlling disclosure. These automated groups can reduce the burden of specifying privacy settings for each individual contact by collating contacts with which information is shared similarly. But even though group formation could be automated, the user would still have to select the appropriate groups for every piece of content that she wishes to share. There remains potential to reduce user burden even further by automatically recommending appropriate groups for sharing a particular piece of content.

Previous studies, e.g. [2, 22], have attempted to identify reasonable predictors of public vs. private settings for sharing content such as photographs, but they have generally not identified predictors that are applicable to various 'non-public' privacy settings, such as restricting access to particular subgroups within a social network. The first contribution of this paper is an investigation of the factors that affect users' decisions to share or not share content with particular contacts from their social network.

A potential barrier for group-based privacy control is that personal privacy policies are often highly dynamic and may vary depending on the current context, need and activity [18]. It may not be feasible to configure groups once in advance and have those groupings hold for all situations. Without consideration of how to share particular instances of content, preconfiguration of groups relies only on information relating to the identity of contacts, available details of their relationships and generalized notions of content sharing from previous experiences. In reality, specific properties of the content being shared may influence the salience and significance of particular group divisions, necessitating the adjustment of group memberships specifically for that context and content. Our second contribution then is an investigation into the contextual dynamics of group membership adjustment. We analyze the efficacy of both static, acontextual grouping and dynamic, contextualized grouping and identify indicators of when each is appropriate.

We conducted a longitudinal study (over 2 months) in which 22 participants used an application with granular access control for sharing photographs with their Facebook contacts. We analyzed the sharing decisions of these 22 participants for a total of 1,014 photos. From this analysis, we quantified some of the factors affecting privacy decisions and assessed whether preconfigured groups for privacy control supported these decisions.

RELATED WORK

Social networking services present many advantages for information dissemination and interpersonal communication, but the copresence of multiple social groups from different facets of a user's life can present a significant challenge for controlling privacy and online

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

identity. Many users experience a perceived loss of control over their personal information and content when using online social networking services [12].

Default privacy settings on services such as Facebook are often configured such that content is shared uniformly with all of a user's contacts. Achieving fine-grained control is an arduous process, yet people consider such control important for presenting multiple versions of themselves [7] or for minimizing the appearance of characteristics that are contrary to an idealized version of themselves [10]. Users often resort to defensive strategies for privacy protection in the absence of easy-to-use granular controls for disclosure [14]; e.g. by not sharing content unless they are comfortable with it being seen by all of their contacts, or by adjusting their content so that it is suitable for a broader audience.

Ackerman and Mainwaring [1] emphasize that, while valued, privacy is not the users' primary task and making it an explicit task for the user can be problematic. Designing privacy management tools that do not require significant configuration effort from the user is therefore an important and worthwhile objective. Systems that automate, recommend or assist with privacy management decisions could reduce the burden placed on users while providing satisfactory levels of control. Raento and Oulasvirta [20] propose a number of design principles for supporting users with privacy management. A main principle is that privacy management tools should allow users to present themselves differently to different audiences. They suggest that this should apply even if disclosure is automated.

In order to inform the design of such tools we need to understand how people make decisions about sharing their content with social network contacts. Lederer et al. [15] found that people base sharing decisions more on the identity of the recipient than on the situation within which the information is sought. Many studies, e.g. [7, 18, 21], have shown that people want to specify groups and categories based on identities and relationships, for which they can specify appropriate disclosure settings. Jones and O'Neill [13] investigated people's rationales when grouping their contacts for the purpose of privacy control, and found six criteria, related to the identities of contacts, which were commonly considered and could inform the automated generation of privacy-based groups.

Although many contextual factors can affect privacy decisions, Lederer [15] suggests that the primary index for such decisions should be the identity of the recipient. The specific context surrounding the disclosure is secondary and has less influence on the decision. However, context cannot be ignored. Altman [3] and Palen and Dourish [19] view privacy as a dialectic and dynamic boundary regulation process, with privacy continuously negotiated and managed and boundaries refined according to context. This highlights a problem for group-based privacy controls: the context in which groups are configured may affect the configuration of group boundaries. Groups based on identity are likely to be fairly stable but not completely unaffected by context. Groups may be different when grouping in advance for 'generic privacy control' and when contemplating disclosure of particular content in the moment.

There has been little research into privacy controls that consider the combination of both salient divisions within a network of contacts, based on identity, and characteristics of relationships and network divisions that become relevant in a given context for sharing a particular piece of content. Gilbert and Karahalios [9] suggest that privacy controls based on tie strength may help to segment a user's social network into meaningful groups. But no work has so far considered the varying importance of tie strength in different contexts. Davis et al. [6] developed MMM2, a system that demonstrated how contextual information, primarily people's copresence, related to a particular piece of content could be used to offer users a list of suggested recipients with whom the user might wish to share. But they did not incorporate information about the identity of the recipients – such as strength and type of relationship – which may also influence the decision to share.

For this paper, we were interested in identifying quantifiable variables relating to both identity of contacts and contextual factors, which facilitate the understanding of an individual's decision to share or not to share a particular photograph with preconfigured sets of contacts. We were also interested in determining an additional set of variables useful for determining whether the use of preconfigured groups is appropriate in a given context and assessing the extent to which preconfigured privacy based contact groups had to be adjusted to be suitable for different contexts.

METHOD

22 participants (11 male, 11 female, mean age 27, range 19-43 years) were recruited through advertisements placed on online notice boards and notices placed around a university campus. Participants were screened through e-mail and face-to-face discussions to ensure that they used their iPhone as their primary camera, they regularly used it to take photographs (typically >20 photos per month), and they often carried it with them wherever they went, enabling photographs to be taken in different contexts.

We also ensured that all of our participants were Facebook users with established networks of contacts and that they were familiar with sharing photographs and other content using the service. Participants were offered a £50 recruitment incentive for completing the study. We did not set a target for the number of photographs they had to take. This reduced the likelihood of participants arbitrarily taking photographs to satisfy the requirements of the study and increased our confidence that they were taking photographs according to their usual reasons for doing so.

Phase 1 – Grouping Exercise

In Phase 1 participants were instructed to create their own privacy-based groups from their Facebook contacts. They could create as many or as few groups as they felt necessary for effectively controlling disclosure of their social media content, with each group containing contacts with whom they would share information/content similarly.

Jones and O'Neill [13] found that people formed privacy groups by creating and combining 'sets' of contacts based on six commonly considered criteria which represented meaningful ways of dividing their social network: Social Circles and Cliques, Tie Strength, Geographical Locations, Organizational Boundaries, Temporal Episodes, and Functional Roles. Our participants were first instructed to identify all of the distinct 'sets' of contacts resulting from dividing their network according to each of these criteria in turn. To assist them, a list of all of their Facebook contacts was provided. Each participant then sorted and aggregated her sets into discrete privacy-based groups of contacts. For example, one participant defined 3 sets of contacts based on the criterion Social Circles & Cliques: 'Family', 'Best Friends', and 'Pub Mates' (in addition to other sets based on other criteria). He then used these 3 sets to define 2 contact groups for content sharing control: Group A – 'Family and Best Friends' and Group B – 'Pub Mates'.

Jones and O'Neill [13] also found that participants typically used Tie Strength (e.g. descriptions of relationship strength) to divide sets of contacts that had been created based on the other criteria, for example splitting a single set into separate 'Strong Tie' and 'Weak Tie' sets. In Phase 1, tie strength divisions were performed last, allowing participants to subdivide any of the sets they had already created if they felt it would help them to control disclosure of content. For example, a participant based 3 contact sets on the criterion Organisational Boundaries: 'Management', 'My Office' and 'My Department'. He then split 'My Office' into 2 sets: 'Strong Ties-My Office' and 'Weak Ties-My Office'. Finally, he used these 4 sets to define 2 privacy-based groups: Group A – 'Management, My Department and Weak Ties-My Office', and Group B – 'Strong Ties-My Office'.

As participants identified the distinct sets of contacts within their network we asked them to consider including contacts that they had deleted, ignored friend requests from or refrained from adding. Users often delete friends as a way of controlling privacy. However, grouping contacts in order to make selective content sharing easier may be preferable to permanent deletion and could be considered a less "anti-social" option.

Phase 2 – Field Study

In Phase 2, the same 22 participants each had an application called 'Flickit' installed on their iPhones for 2 months. They were asked to use the Flickit application instead of the default iPhone camera application any time they wanted to take a photograph. They positioned the Flickit icon on the first screen of their iPhone, next to the default camera application so that they would be reminded to use it.

Figure 1. Providing photo meta-data and choosing sets of contacts for sharing this particular photo

Flickit is intended for managing the upload of photographs to different albums on Flickr. We used the application to replicate managing the disclosure of content with social network contacts, by mapping each of the sets that participants had created in Phase 1 to a Flickr album, using the set labels as album names.

Each time they took a photograph, users were presented with a list of their sets of contacts (Fig. 1). Participants could then decide with which sets to share the photograph. This grounded sharing decisions in a particular context, where different factors could be salient for managing group boundaries. In order to test the efficacy of preconfigured groups for managing privacy control, only the sets, rather than the groups, from Phase 1 were embedded in the interface. The privacy groups were created in Phase 1 as collated sets of contacts with which information would be shared similarly. If all of the sets that comprised a Phase 1 privacy-based group received the same sharing decision for a given photo in Phase 2, then we can infer that that group was suitably preconfigured for use in that specific context. If, however, the decision to share or not to share was not applied uniformly across different sets of contacts from within the same privacy-based group, then that group did not serve its purpose of collating contacts with whom the participant would share the same information.

Once sharing decisions had been made for all sets of contacts they were uploaded to private accounts on Flickr so that we could access them for analysis. To guard against inadvertent privacy violations, photos were not actually shared with our participants' contacts or uploaded to Facebook, but participants were instructed to treat their use of the system as if they would be, or could be in the future.

In order to reduce the risk of bias as a result of participants not uploading photos that they did not want to share with the researchers, we provided the option to upload a blank replacement image for the photograph that they had taken. In this case, we asked them to provide information about

the image as if it were the photo they had taken. This method allowed us to record participants' sharing decisions without having to reveal their more sensitive photos.

CONTENT AND CONTEXT METADATA

In addition to selecting sets of contacts with whom to share photographs, participants provided related details for analysis. Flickit allowed pre-configured tags to be assigned to images, which we used to collect additional data. Our metadata collection was informed by prior research, e.g. [2, 6, 23], recording data on elements of context and content that have been shown to affect privacy decisions: Location, Event, Copresence, Content and Motivation.

We reason that these could be useful as explanatory variables for understanding sharing decisions, for use within development of recommender systems for content sharing and also for analyzing the contextual dynamics of group formation/adjustment.

Location is often indicative of significant contexts in peoples' lives. Ahern et al. [2] found that a significant portion of users have some locations in which they are more likely to take private photos, and some in which they are more likely to take public photos. We incorporated location data collection in order to investigate whether location was also useful for understanding decisions to share with particular sets of contacts, beyond the simple public vs. private distinction.

Bentley and Metcalf [4] found that events are a meaningful concept for organizing photos, and Lovett et al. [16] suggest that they may be useful for interpreting individual and group context. Events can be segmented by both location and time. We collected location and time data that had been automatically added to the photo's EXIF data by the application. To enable us to make comparisons between different types of events/locations both within and between participants, we also asked them to specify tags describing the type of event and place at which the photo was taken.

Previous work, e.g. [6], has shown that detecting people who are copresent when photographs are taken is useful for suggesting recipients. People often like to share photographs with others who have shared an experience and may have fewer privacy concerns about revealing such photos to people who were present. We therefore asked our users to give tags describing whom they were with when they took the photo.

Ahern et al. [2] found that users have content-derived patterns in making privacy decisions. We asked participants to specify tags which would help to describe the content of the image according to categories of content identified in [2]: People, Place and Object. We added an Animal category after piloting the study and finding that animals were frequently the subjects of photographs.

Some research has suggested that privacy concerns may vary according to the motivations for taking photos. For example, Miller and Edwards [17] found that for people who use photos to capture memories about which stories can later be told, privacy was the most important factor in determining whether to share photos. For others who see photography as a way to reflect on even the most mundane experiences, privacy was less of a concern. We attempted to capture people's motivations for taking photos by asking them to specify a tag from a taxonomy of motivations identified in [23]: creating and maintaining social relationships; constructing personal and group memory; self-presentation; and self-expression and functional use. Table 1 shows the tags categorized by element of context.

Category	Tags
Event/Location	*home, work, school, party, pub, nightclub, commuting*
Copresence	*alone, with friends, with family, with colleagues, with strangers*
Content	*photo of object, photo of person, photo of place, photo of animal*
Motivation	*social relationships, capture memory, self-expression, functional use*

Table 1. Tags and Content Metadata

The complete set of tags was presented to the participants for each photo that they took. They were instructed to select all applicable tags for each photo, with the option to add their own additional tags if necessary. Participants were also given a 'Description' field (Fig. 1), in which they could enter comments about their sharing decisions. If a participant tried to share a photograph and found that the sets that he had created did not provide adequate control over disclosure, he was instructed to provide a brief explanation of why this was the case, for example, if there were particular individuals whom he decided to include or exclude, or if the contacts that he wanted to share with had not been defined as a set.

Phase 3 – Post-Study Questionnaire and Interviews

Following the field study, all participants completed a questionnaire in order for us to gain further insight into how they had made sharing decisions. We followed up with individual interview sessions for the 10 participants we were able to meet face to face. The questionnaire and interviews explored the needs and practices of sharing information and examined how well the access control mechanism of the photo sharing application supported user requirements. We use some of our qualitative data to offer explanations for our quantitative findings, however a full description of our findings is beyond the scope of this paper.

ACONTEXTUAL GROUPING EXERCISE RESULTS

On average, participants used the 6 criteria to divide their contacts into 18 distinct sets (S.D.=3.85), which they then used to create a mean of 3.8 distinct privacy groups (S.D.=1.66). Fig. 2 shows a distance matrix for contacts associated with the different criteria. Distance was calculated using the frequency at which sets associated with

each criterion appeared in the same privacy group, i.e. the number of times a set from criterion A was grouped with a set from criterion B. We combined these frequency measures for all groups created by all participants and calculated the average frequency to produce a complete distance matrix. Black cells (in the 75-100% region of the frequency range) indicate a strong propensity for participants to group together contacts associated with the corresponding criteria. Light and dark grey cells (in the 25-50% and 50-75% regions of the frequency range, respectively) indicate a lower frequency of contacts with the corresponding criteria being grouped together. White cells (0-25%) indicate a very low frequency of contacts grouped together with these criteria.

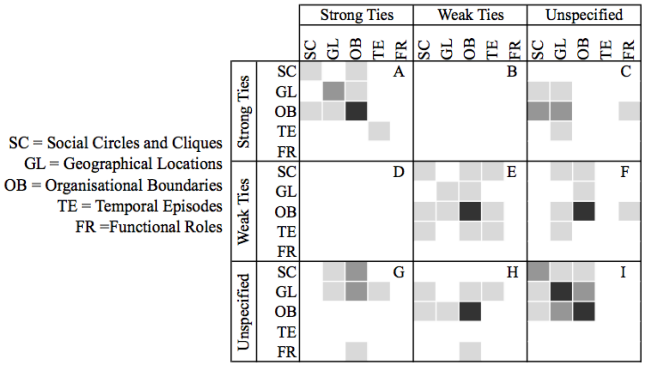

Figure 2. Distance matrix for contact sets

Figure 2 shows that there was a tendency not to mix different levels of tie strength within privacy groups. Strong ties were often grouped with other strong ties (Region A) and weak ties were often grouped with other weak ties (Region E). Contacts of different tie strengths were never combined, regardless of what other criteria they may have had in common (Regions D and B).

These results also reveal the level of granularity required in defining groups. For example, contact sets from within organizational boundaries were often merged with contact sets from within other organizational boundaries of similar tie strength to form a single group for content sharing purposes. Hence, there may be no need to separate contacts from within such contexts with fine granularity; it would suffice simply to aggregate all contacts defined by the Organizational Boundary criterion, whatever organization and therefore set they belong to, and split this 'superset' based on tie strength alone. This was also the case for Geographical Locations, albeit to a slightly lesser extent.

The same was not true for other criteria. For example, contact sets based on functional roles (relationships which are not defined by 'friendship' but some other functional reason for interacting with one another) were very rarely grouped with sets based on any other criteria, or even with contacts based on the same criterion. This indicates the need for fine granular control of these particular contacts, and highlights that different granularities are required for different relationship types within a social network.

FIELD STUDY RESULTS

In total we recorded 18,287 decisions to share or not to share a photo across all participants. We used these data to produce an analytical model for sharing decisions using logistic regression. The model provides us with a formula that computes the probability of a sharing decision as a function of the values of a number of categorical explanatory variables, in this case the metadata associated with each photograph. We were interested in identifying a set of quantifiable variables that will be useful in predictive models for selecting or recommending an appropriate audience for sharing.

Binary Logistic Regression Analysis

We conducted a hierarchical logistic regression on our dataset of sharing decisions. For our response variable we gave 'Share' responses a value of 1 and 'Do not share' a value of 0. For our binary explanatory variables we were able to use our automatically collected metadata and user selected tags relating to location, event, content, copresence, motive etc, as well as information relating to the contact's identity, such as the strength of ties and the set criteria with which the contact was associated. Incrementally adding blocks of variables to the model allowed us to examine whether the newly incorporated variables provided improved prediction ability over the preceding model; i.e. each block accounts for a certain amount of the variability observed in our 'Share Decision' measure.

Figure 3 shows the parameters for the logistic regression and our resultant analytical model of sharing decisions. B is the estimated coefficient for each variable in the model's equation. If the Wald statistic is significant ($p < 0.05$) then the parameter is useful to the model. The odds ratio; the predicted change in odds for a unit increase in the binary explanatory variable (i.e. from 0 to 1) can be calculated using Exp(B).

In the first stage of our logistic regression analysis, we examined individual differences between our 22 participants (using 21 orthogonal contrasts). We found that different participants had different propensities to share photographs, i.e. some of our participants were more inclined to share than others. We hypothesized that these differences might be associated with the participants' privacy concerns, however a test of Pearson's correlation coefficient did not reveal significant associations between coefficient/odds ratio and scores for a privacy concern questionnaire taken from [5] which participants completed at the end of Phase 1 ($r = 0.149$, $n=22$, $p=0.22$, n.s.).

In the second stage of our logistic regression analysis we examined the relationship between contacts' tie strength and decisions to share with them. We found that identifying both strong tie ($B=0.78$, $p<0.05$) and weak tie ($B= -1.328$, $p<0.05$) contacts offered significant influence within our model. Participants were twice as likely to share photos

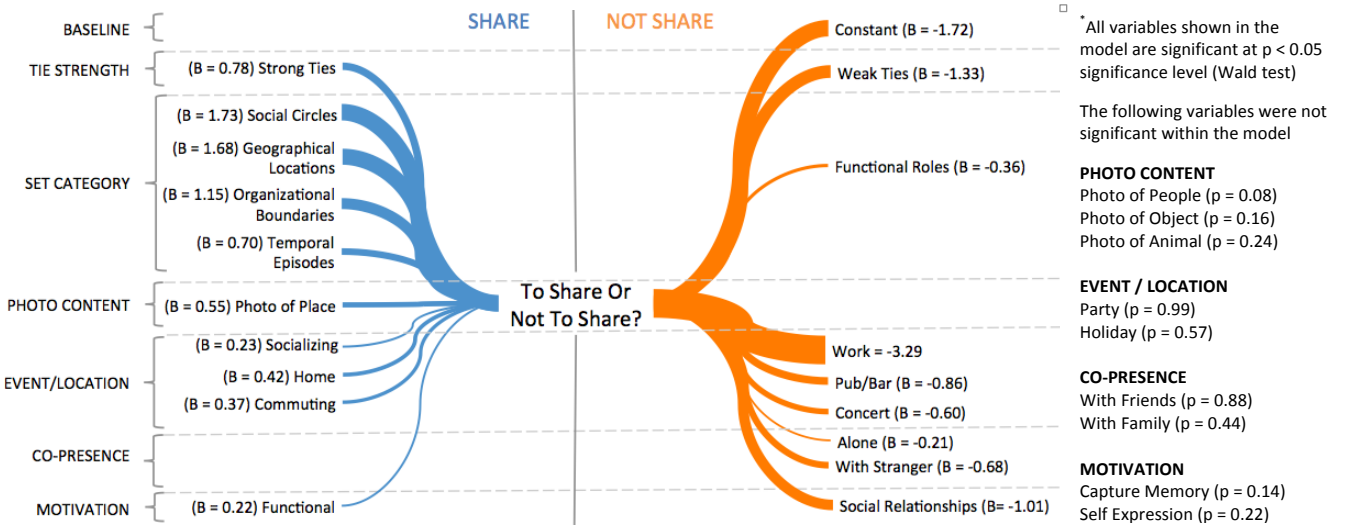

Figure 3. To Share Or Not To Share Logistic Regression Model

with contacts if they were classified as strong ties (compared to not strong ties). Weak ties were correlated with the decision not to share. Participants reported that they were often concerned about the implications of sharing personal content with people whom they did not know well, but were more comfortable sharing content with closer contacts.

In the third stage we examined whether the criteria for dividing network contacts affected sharing decisions. We found that participants were most likely to share with sets of contacts that corresponded to social circles/cliques in their network (B=1.733, p<0.05) and geographical locations (B=1.681, p<0.05), followed closely by organizational boundaries (B=1.151, p<0.05) and then temporal episodes (B=0.701, p<0.05). Only sets corresponding to functional roles (B= -0.355, p=0.05) decreased the odds of sharing.

In the fourth stage we examined the effects of photo content. Our model indicated that identifying whether or not a photograph contained people was not significantly useful for predicting sharing decisions (B=0.184, p=0.08). Results from our post-study questionnaire and interviews suggested, however, that many of our participants had considered whether there were people featured within the picture when making their sharing decisions. We discuss this subject further in the Discussion section below. Similarly, identifying that an object (B=0.131, p=0.16) or an animal (B=0.126, p=0.24) was the subject of a photograph did not offer significant predictive ability.

We found that photos of places, rather than people, objects or animals, were useful for predicting sharing decisions. This content increased the odds of sharing (B=1.731, p<0.05). Many of our participants commented that it was often 'easy' to make sharing decisions when the photos were of places. Generally they were more willing to share such photos due to the perception that they did not reveal personal information.

In the fifth stage we examined events and locations. Our model indicated that there was a strong decrease in the odds of sharing a photograph if it was taken at work (B= -3.292, p<0.05). Perhaps surprisingly, photographs taken at a bar/pub (B= -0.856, p<0.05), party (B= -0.004, p<0.05) and concert (B= -0.599, p<0.05) also had an overall effect of decreasing the likelihood of the photo being shared. Our participants offered a possible explanation for this observation in their interviews, commonly stating that while they took many photographs in these settings, they were the most likely to show them in embarrassing or compromising situations. Photos tagged as related to 'Holiday' did not provide statistical significance within the model (B=0.062 p=0.57). This suggests that participants considered other aspects of context as more important when deciding with whom to share holiday photos.

In the sixth stage we examined how the presence of others at the time of photo capture affected sharing decisions. Both being alone (B= -0.214, p<0.05) and the presence of strangers (B= -0.681, p<0.05) were negatively correlated with photographs being shared. The presence of friends (B= -0.013, p=0.88) and family (B=0.075, p=0.44) were not significant regressors within the model.

In the seventh stage of our logistic regression analysis we incorporated explanatory variables relating to the motivation for taking a photograph. Our analysis revealed that photographs taken to capture social relationships between people were less likely to be shared (B= -1.011, p<0.05). During our interviews it was common for participants to discuss their reluctance to share what they deemed to be photos of a more personal nature and these were often the photographs that they had taken to capture relationships with friends, family and loved ones.

Photographs that were taken to capture personal and group memories of events (B= -0.112, p=0.14) and those which were used as a form of self expression (B=0.109, p=0.22)

were not significant predictors of sharing decisions. We suggest that the motivation 'capturing personal and group memories' is too vague and encompasses such a wide range of contexts and content that it does not offer any predictive power. Many of the photos that were uploaded with the 'self expression' tag seemed to exemplify more traditional artistic and aesthetic goals of photography. We found that our participants had difficulty knowing when to assign this tag, as it was not always clear to them what constituted 'self expression'.

Our model also includes a significant constant component (B= -1.723, $p<0.05$), indicating that our participants exhibited a slight inclination not to share content unless other variables motivated them to do so.

In order to verify that the model accurately reflects sharing decisions made by our participants it was important to assess how well the model fitted our collected data.

Model Classification Measures

We assessed the fit of our classification model by examining the percentages of sharing occurrences correctly classified (Sensitivity), non-occurrences correctly classified (Specificity), sharing occurrences that were incorrect (False Positive Rate) and non-occurrences that were incorrect (False Negative Rate).

In total we included 18,287 sharing decisions (to share or not to share a particular photo with a particular set of contacts) in our analysis. 79.3% of our model's classifications were correct, indicating a reasonable fit to our data. (Sensitivity: 86.8%, Specificity: 68.7%, False Positive Rate: 21.4%, False Negative Rate: 20.3%). Our model offers a significant reduction in the number of misclassifications over the constant model, which simply shares with everyone and achieves only 50.4% accuracy.

We emphasize that we do not see our model as a ready-to-use solution for building a prediction/recommender system. Rigorous cross-validation and testing is necessary to assess how a predictive tool would work in the real world. It is still unknown precisely what level of accuracy a real recommender system would require; anything significantly better than chance might be of some use. A recommender system could engage in some confirmatory dialogue with the user or allow her to refine the recommendations, providing control while still reducing the burden of manually selecting an audience. Table 2 shows how our model classified the sharing decisions.

Observed		Predicted		
		Sharing Decision		Percentage Correct
		Not Share	Share	
Sharing Decision	Not Share	9310	1413	86.8
	Share	2364	5199	68.7
Overall Percentage				79.3

Table 2. Predicted vs. Observed Sharing Decisions

Our R^2-type statistics (Cox & Snell $R^2 = 0.467$, Nagelkerke $R^2 = 0.594$) indicate that the included explanatory variables capture a considerable amount of the data variability.

In addition to creating a single classification model for all of our participants, we used the same process to generate models for each individual participant. Our results suggest that by fitting different models to individuals based only on their own sharing decisions, the fit to our real data could be improved. Misclassifications were slightly reduced, with an average of 84.2% (S.D.=3.9%) of sharing decisions being correctly classified across the 22 individual models.

In the next section we tackle an issue that may offer the ability to identify situations in which it is not feasible to predict the outcomes of group-based privacy decisions. We examine the occurrence of situations in which sharing decisions do not align with participants' preconfigured privacy groups from Phase 1, i.e. the contexts in which these preconfigured groups needed to be significantly adjusted to provide the correct group for sharing a particular item.

GROUP ADJUSTMENT RESULTS

For each photograph we identified the contact sets, chosen via the application, with which the participant had decided to share that photo. We then calculated the extent to which the sets within each of the Phase 1 privacy-based groups received the same sharing decision. This provided us with a measure that we refer to as group cohesiveness; in other words, the degree of similarity between actual disclosure groups and preconfigured groups. Groups that have high cohesiveness are rarely adjusted, whereas groups with low cohesiveness are frequently reconfigured, perhaps according to considerations that become significant within a particular context.

For all 1,014 photos we found that, on average, 90.6% of sets within the same Phase 1 privacy group received the same disclosure settings (SD=0.09%). This suggests that the Phase 1 privacy-based groups created by participants were reasonably effective at grouping contacts with whom information would be shared similarly but that some adjustment was required. Allowing participants to specify sharing decisions at a 'Group' granularity would most often result in the correct decisions being applied to the contained sets of contacts, however, it is important that users are able to adjust groups when necessary as even a single unwanted recipient of sensitive content might nullify the benefits of the group-based approach.

We conducted a similar analysis with logistic regression, this time using group adjustment as our binary response variable. When all sets in a group were shared with uniformly, the group received a cohesiveness score of 1. If groups were adjusted, i.e. some member sets received different sharing decisions to other sets in the group, then the cohesiveness score was calculated as the largest fraction of the group to which the same setting was applied.

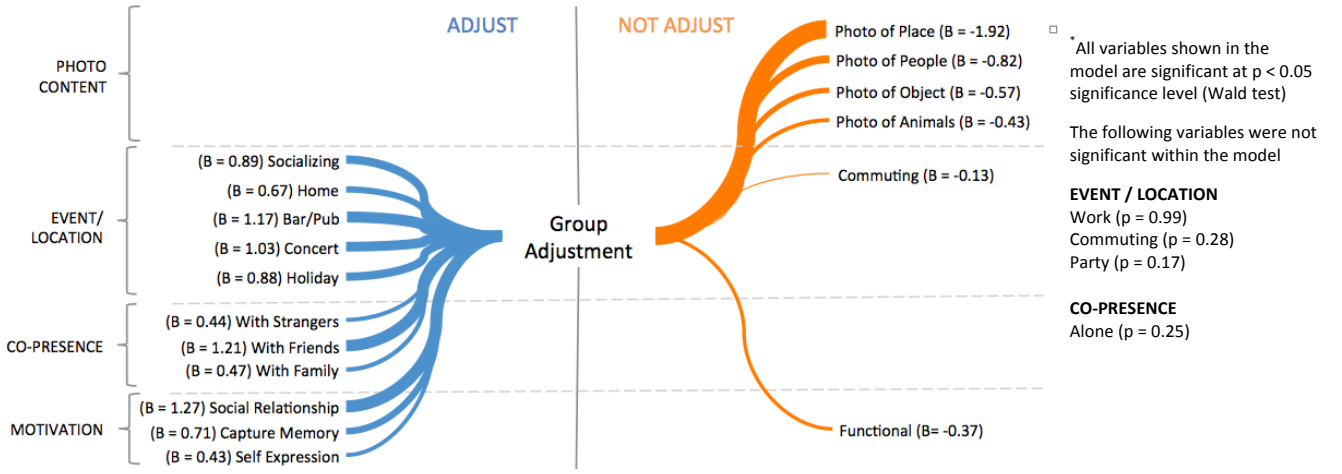

Figure 4. Group Adjustment Logistic Regression Model

We were interested to see if we could use the contextual information provided with the photographs to identify when group adjustment was most likely to take place. As this was an exploratory analysis, we set a reasonably low threshold of 0.9 for our group cohesiveness measure (i.e. 90% similarity between a Phase 1 privacy based group and the corresponding context specific sharing sets), below which we classified photos as necessitating the adjustment of preconfigured groups. It may be the case that even the slightest of adjustments become important for avoiding privacy violations. However, we were first interested to see if particular contexts required large adjustments.

Figure 4 shows the results of our logistic regression analysis and the coefficients in the model for the occurrence of group adjustment. We used the model to make statistical inferences about the contexts in which group adjustment was most marked. As with our previous model, we assessed its fit to our data, finding that the model offered a significant improvement over the naïve constant model (80.6% correct classifications). For 54.7% of sharing decisions no group adjustment took place. If groups were adjusted, on average it was by adding or removing no more than 9.4% of its sets.

A breakdown for our model is – Sensitivity: 88.0%, Specificity: 67.1%, False Positive Rate: 16.9%, False Negative Rate: 24.7%. Given the model's fit to our data, the results from the model provide reasonable predictors of group adjustment.

In the first stage of our logistic regression analysis, we examined individual differences and found that there were no significant differences in the occurrence of group adjustments between any of our participants. This implies that our participants had similar approaches to group adjustment.

In the second stage of our logistic regression analysis, we examined whether particular content was useful for predicting group adjustment. We found that all of our content related explanatory variables were useful for predicting whether groups would be adjusted ($p<0.05$). We found that photographs of people ($B= -0.822$, $p<0.05$), objects ($B= -0.565$, $p<0.05$) and animals ($B= -0.426$, $p<0.05$) had similar negative correlations with group adjustment. Photographs of places had a smaller, but still negative, effect on the odds of groups being adjusted ($B= -1.915$, $p<0.05$).

In the third stage of our logistic regression analysis, we examined events and locations. Among the remaining variables we found that neither work ($p=0.99$), 'commuting' ($B= -0.129$, $p=0.28$), nor 'party' ($B=0.339$, $p<0.17$) were significant regressors. All other variables were significant within the model and increased the odds of groups being adjusted: socializing ($B=0.885$, $p<0.05$), home ($B=0.673$, $p<0.05$), pub ($B=1.173$, $p<0.05$), concert ($B=1.033$, $p<0.05$), and holiday ($B=0.878$, $p<0.05$).

In the fourth stage of our logistic regression analysis, we examined copresence. Interestingly, we found that photos that were taken in the presence of others were correlated with privacy-based groups being significantly adjusted; this was most notable when photos were taken in the presence of friends ($B=1.206$, $p<0.05$). Being in the presence of strangers had a weaker positive correlation with groups being adjusted ($B=0.441$, $p<0.05$). Being alone was not a statistically significant predictor of group adjustment ($B= -0.163$, $p=0.25$). In the fifth stage of our logistic regression we incorporated explanatory variables relating to motivation. We found that photos for creating and maintaining social relationships were positively correlated with the adjustment of groups ($B=1.270$, $p<0.05$). The same was true for photos capturing personal and group memories ($B=0.713$, $p<0.05$) and self expression ($B=0.434$, $p<0.05$). Photos taken for functional purposes were negatively correlated with group adjustment ($B= -0.366$, $p<0.05$).

DISCUSSION

Our analysis revealed useful explanatory variables relating to the identity of contacts and the context of photo capture: namely, where and when the photo was taken, who was present at the time, what the photo was of, who was in the photo, why it was taken, the strength of the relationship to a potential recipient and the salient features of that recipient's identity with respect to divisions in the sharer's social network.

Our models offer a range of insights that may be useful to social network service designers, e.g. people are more reluctant to share photos capturing social relationships than photos taken for functional purposes; certain settings such as work, bars, concerts cause users to share less; and being a weak tie decreases the chance of being shared with in a system with granular access control.

Systems that aim to automate the process of selecting audiences for content sharing should incorporate important variables that we have identified within our models. Our findings also present design implications with respect to the configuration and adjustment of groups. Groups do not often incorporate both weak and strong ties, and fine levels of granularity are not required within groups based on organizational boundaries. Interfaces for creating groups could increase the efficiency of organizing contacts by accounting for these findings. With regard to group adjustment, our findings illustrate that systems providing group based access control should provide the ability to adjust groups after initial configuration. The frequency at which groups need to be adjusted in order to facilitate the needs of users and the contextual dynamics of sharing decisions emphasizes the importance of making this an efficient process.

Feedback from our participants in questionnaires and interviews revealed that the main criticism of the photo sharing application that they used was the effort required manually to specify granular privacy settings on a per photo basis, yet they valued the control it gave them over their privacy. Participants also found it time consuming to provide photo metadata, which was useful for our analysis. Although some of this information was automatically captured without explicit input from the participant, many features of the content and context had to be described manually by selecting appropriate tags. There are substantial bodies of work that aim to capture and describe the context of mobile devices and their users, as well as work that aims to automatically identify features of photo content. We expect that as these fields progress, our reliance on users to provide such information to predictive tools will be significantly reduced. We cannot know the extent to which machine learning will obviate the need for explicit input. Our study instead focuses on highlighting which factors matter for organizing future studies.

We found that some variables were not significantly useful for predicting sharing decisions, however a possible limitation of our approach is that these variables were not captured at an appropriate resolution. For example, we found that determining whether people are featured in a photograph was not particularly useful. But in our questionnaires and interviews, participants frequently mentioned that precisely *who* appeared in the photo was important, rather than the fact that *somebody* appeared in the photo. There are also additional variables that we were not able to control for but which participants suggested might have some influence on sharing decisions. For example, the strength of their relationship with the person in the photo mattered, as this sometimes caused them to consider not only their own privacy but also the privacy of other people in the photograph.

A possible limitation of our predictive model is that we do not consider differences between sharing decisions that are made 'in the moment', i.e. immediately after the photo has been taken, and those that are made some considerable time later. Users are sometimes uncertain about the content of, audience for, and norms regarding particular disclosures, and this uncertainty limits their ability to make the best decision at capture time [2]. Our model aims to replicate sharing decisions made by the users, with the assumption that they do not lead to privacy violations, although this may not always be true. Some of our findings suggest that users may have taken less time and/or care over privacy decisions when they were busy; e.g. photos were shared less often when users were with friends and the decisions they made correlated less with the privacy groups that they had preconfigured. Future work should look to gain further insights into the causes of these observations.

Although our study focused solely on photo sharing with social network contacts, we present our methodology as a useful outline for investigating patterns of disclosure with other types of social media content. For example, disclosure decisions for status updates may vary according to the people or subjects they reference, locations they relate to, strengths and types of relationships with potential recipients, and so on. Many of our participants told us that they would appreciate a similar method of access control for their status updates. Other privacy sensitive items of content, such as location sharing updates, e.g. using Facebook Places, create intense debate with respect to user's privacy. Assistive privacy management tools that account for the contextual dynamics of users' sharing decisions provide an opportunity to significantly reduce the burden on the user.

CONCLUSION

Our research addresses the considerable burden of exercising fine-grained control for sharing content with social network contacts. We have presented a novel approach to understanding relationships between properties of SN contacts, content, context and sharing decisions. Our work underlines the complexity of variables affecting sharing decisions but presents a manageable approach to

quantifying their effects and takes a step towards better recommendation systems. We have shown that preconfigured privacy based groups which are intended to simplify and reduce the burden of controlling disclosure with individual contacts do not always provide a suitable mechanism for sharing in *all* contexts, as the groups often have to be adjusted. Furthermore, we have identified some factors that affect users' tendency to adjust these groups.

We expect that social network users will benefit from the ability to easily to share content with different people based on both their differing relationships and the specific context relating to that content. Our future work will investigate and test predictive content sharing tools and recommender systems based on our models in order to improve users' content sharing experiences.

ACKNOWLEDGMENTS

Eamonn O'Neill's research is supported by a Royal Society Industry Fellowship at Vodafone Group R&D.

REFERENCES

1. Ackerman, M. and Mainwaring, S. (2005). Privacy Issues in Human-Computer Interaction. In L. Cranor and S. Garfinkel (Eds.), Security and Usability: Designing Secure Systems that People Can Use, 381-400, Sebastopol, CA, O'Reilly.
2. Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M. and Nair, R. 2007. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. Proc. CHI '07. ACM, 357-366.
3. Altman, I. (1975). The Environment and Social Behavior: Privacy, Personal Space, Territory and Crowding. Monterey, CA: Brooks/Cole Pub. Co., Inc.
4. Bentley, F. and Metcalf, C. (2006). Flexible Views: Annotating and finding context-tagged mobile content. Ubicomp 2006 workshop on Pervasive Image Capture and Sharing.
5. Buchanan, T., Paine, C.B., Joinson, A.N. and Reips, U-D. (2007). Development of measures of online privacy concern and protection for use on the Internet, Journ. of the American Society for Information Science and Technology 58. 157–165.
6. Davis, M., Canny, J., Van House, N., Good, N., King, S., Nair, R., Burgener, C., Rinehart, B., Strickland, R., Campbell, G., Fisher, S. and Reid, N. (2005). MMM2: mobile media metadata for media sharing. Proc. MULTIMEDIA '05. ACM, 267-268.
7. DiMicco, J. M. and Millen, D. R. (2007). Identity management: multiple presentations of self in facebook. Proc. GROUP '07. ACM, 383-386.
8. Fang, L. and LeFevre, K. (2010). Privacy wizards for social networking sites. Proc. World Wide Web (WWW '10). ACM, New York, NY, USA, 351-360.
9. Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. CHI '09. ACM, 211-220.
10. Goffman, E. (1959). The Presentation of Self in Everyday Life. New York: Doubleday.
11. Granovetter, M. S. (1973). The Strength of Weak Ties. The American Journal of Sociology, vol. 78, no. 6, 1360–1380.
12. Hewitt, A. and Forte, A. (2006), Crossing boundaries: Identity management and student/faculty relationships on the Facebook, Proc. CSCW06. ACM.
13. Jones, S. and O'Neill, E. (2010). Feasibility of structural network clustering for group-based privacy control in social networks. SOUPS '10, vol. 485. ACM, 1-13.
14. Lampinen, A., Tamminen, S. and Oulasvirta, A. (2009). All My People Right Here, Right Now: management of group copresence on a social networking site. GROUP '09. ACM, 281-290.
15. Lederer, S. Dey, A. K. and Mankoff, J. (2002). A conceptual model and a metaphor of everyday privacy in ubiquitous computing, Intel Research, Tech. Rep. IRB-TR-02- 017.
16. Lovett, T., O'Neill, E., Pollington, D. and Irwin, J. (2009) Event-based mobile social network services, in Workshop on Context-Aware Mobile Media and Mobile Social Networks, Mobile HCI 2009.
17. Miller, A. D. and Edwards, W. K. (2007). Give and take: a study of consumer photo-sharing culture and practice. Proc. CHI '07. ACM, 347-356.
18. Olson, J., Grudin, J. and Horvitz, E. (2005). A study of preferences for sharing and privacy. CHI '05 extended abstracts. ACM, 1985–1988.
19. Palen, L. and Dourish, P. (2003). Unpacking "privacy" for a networked world. Proc. CHI '03. ACM, 129-136.
20. Raento, M. and Oulasvirta, A. (2008). Designing for privacy and self-presentation in social awareness. Personal Ubiq. Computing. 12, 7 (Oct. 2008), 527- 542.
21. Skeels, M. and Grudin, J. (2009). When social networks cross boundaries: a case study of workplace use of facebook and linkedin. Proc. GROUP '09. ACM, 95-10.
22. Tufekci, Z. (2008). Can You See Me Now? Audience and Disclosure Management in Online Social Network Sites. *Bulletin of Science and Technology Studies*. Volume 11, Number 4, June 2008 , pp. 544-564(21).
23. Van House, N., Davis, M., Ames, M., Finn, M. and Viswanathan, V. (2005). The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing. CHI '05 Extended Abstracts. ACM, 1853-1856.